Dis 13 Estimation

Thursday, 22 April 2021 11:04 AM

Agenda

1. Inference overview: hypothesis testing vs estimation
2. Finding MMSE, LLSE (and QLSE)
3. Geometric interpretation and definition of LLSE (projection property)
4. Questions + JG RVs recap

obs grounded truth / cause

$Y \to X$

inference fn

$\hat{X}(\,)$ good!

	X	Quality of 'inference function'	Example
Hypothesis testing	(binary) $\in \{0, 1\}$ discrete	Neyman Pearson formulation	Symptoms → Cancer / No Cancer
Estimation	$\in R$	For 126, we look at MSE (mean squared error)	GPS coordinates from satellite → real position

Definition of MMSE / LLSE / QLSE

$Y \to X$

$\hat{X}(\cdot)$ lower MSE $E[(\hat{X}(Y)-X)^2]$

could have other 'loss' fncs 'L1' $E[\,|\hat{X}(Y)-X|\,]$

$MMSE[X|Y]$ = is fn min. MSE

$= E[X|Y]$ (HW)

$L[X|Y]$ = is linear fn min MSE

$(\equiv a + bY)$

$Q[X|Y]$ quad fn " "

$(\equiv a + bY + cY^2)$

Q: When is LLSE[X|Y] == MMSE[X|Y]? when $E[X|Y]$ is linear

1. Balls in Bins Estimation

We throw $n \geq 1$ balls into $m \geq 2$ bins. Let X and Y represent the number of balls that land in bin 1 and 2 respectively.

(a) Calculate $\mathbb{E}[Y \mid X]$.

(b) What is $L[Y \mid X]$? No calculations needed here.

(c) Suppose we want to calculate $L[Y \mid X]$ using the formula

$$\boxed{L[Y \mid X] = \mathbb{E}[Y] + \frac{\mathrm{cov}(X,Y)}{\mathrm{var}(X)}(X - \mathbb{E}[X]).}$$

def of $L(X|Y)$

by def^n

2 proofs
1) Geometric
2) algebraic

First compute $\mathbb{E}[X]$ and $\mathbb{E}[Y]$.

(d) Compute $\mathrm{var}(X)$.

(e) Compute $\mathrm{cov}(X,Y)$.

(f) Compute $L[Y \mid X]$ using the formula. Ensure that your answer is the same as your answer to part (b).

Now we can derive $L[X|Y]$ geometrically. Our basis for the observation space Y is $\{1, Y\}$, where Y may have nonzero mean. This means that the vector representation of 1 may not be orthogonal to Y. We would like to project X onto this basis, but first it must be orthogonal! This is where Gram-Schmidt comes into play. 1 as a vector is already a unit vector. Our orthonormal Y becomes

$$\tilde{Y} = \frac{Y - \langle Y, 1\rangle}{\|Y - \langle Y, 1\rangle\|}$$

So $\{\tilde{Y}, 1\}$ is an orthonormal basis, so then

$$\begin{aligned} P_{\{1,\tilde{Y}\}}(X) &= P_{\{\tilde{Y}\}}(X) + P_{\{1\}}(X) \\ &= \langle X, 1\rangle \cdot 1 + \langle X, \tilde{Y}\rangle \cdot \tilde{Y} \\ &= \mathbf{E}[X] + \mathbf{E}\left[\frac{X(Y - \mathbf{E}[Y])}{\sqrt{\mathbf{Var}(Y)}}\right]\frac{(Y - \mathbf{E}[Y])}{\sqrt{\mathbf{Var}(Y)}} \\ \implies L[X|Y] &= \mathbf{E}[X] + \frac{\mathbf{Cov}(X,Y)}{\mathbf{Var}(Y)}(Y - \mathbf{E}[Y]) \end{aligned}$$

I'll get back

MMSE$(Y|X)$

$n - x$ balls left

linear

a) $E[Y \mid X = x] = \dfrac{n - x}{m - 1} = \dfrac{n}{m-1} - \dfrac{1}{m-1}X$

$Y|X \sim \mathrm{Bin}\left(n - X, \frac{1}{m-1}\right)$

b) $E(Y|X)$

c) $X \sim \mathrm{Bin}\left(n, \frac{1}{m}\right)$

same for Y

d) $np(1-p)$

e) $X = X_1 + \dots + X_n$, $Y = Y_1 + \dots + Y_n$, Y_i, X_i are indicator for i^{th} ball

$\mathrm{Cov}(X, Y) = \mathrm{Cov}\left(\sum_i X_i, \sum_j Y_j\right) = \sum_i \sum_j \mathrm{Cov}(X_i, Y_j)$

(bilinearity of covariance)

if $X_i = 1 \Rightarrow Y_j = 0$

$= \sum_i \sum_j E(X_i Y_j) - E(X_i)E(Y_j)$

$\mathrm{Cov}(X_i, Y_j)$ ∵ X_i, Y_j are indep

$= \sum_{i \neq j} \left(E(X_i Y_j) - E(X_i)E(Y_j)\right) = 0$

$+ \sum_{i=1}^{n} E(X_i Y_i) - E(X_i)E(Y_i)$, $E(X_i Y_i) = 0$

$n(n-1)\,0$

$= \cancel{n(n-1)\left[\Pr(X_i = 1, Y_j = 1) - \frac{1}{m}\right]}$

$= n \times \left(0 - \frac{1}{m} \times \frac{1}{m}\right) = -\dfrac{n}{m^2}$

Geometric interpretation

Assume X, Y are zero-mean RVs with finite second moments. Then we have the following association of RVs as a vector space:

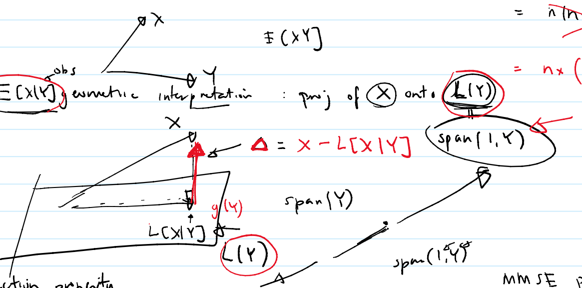
RV X	Geometry
random variable X	a vector
RVs X and Y	two vectors with an angle θ between them
$\mathbf{E}[XY]$	$\langle X, Y\rangle = \|X\|\|Y\|\cos(\theta)$
$\mathbf{E}[XY] = 0$	$\theta = \frac{\pi}{2}$
$\mathbf{E}[X^2]$	$\langle X, X\rangle = \|x\|^2$ which is the norm of X
$\rho = \frac{\mathbf{E}[XY]}{\sqrt{\mathbf{E}[X^2]}\sqrt{\mathbf{E}[Y^2]}}$	$\frac{\langle X,Y\rangle}{\|X\|\|Y\|} = \cos(\theta)$

Table 1: Correspondences between random variables and their geometry

$E(XY)$

LLSE$[X|Y]$ obs geometric interpretation : proj of X onto $L(Y)$

span$(1, Y)$



LLSE$[X|Y] \equiv$ proj X onto $L(Y)$

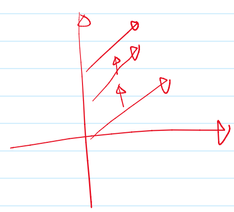
span$(Y) = \{bY, \forall b\}$

$= \{a + bY, \forall a, b \in \mathbb{R}\}$

$=$ span$(1, Y)$ span(Y)

span(P, Q)

$= \{aP + bQ, \forall a, b \in \mathbb{R}\}$



Projection property

if $X - g(Y) \perp$ all elements of $L(Y)$

$\Rightarrow g(Y) = L(X|Y)$

MMSE proj X on $\phi(Y)$

any fnc of Y

orthogonal

span$(1, Y)$

$X - L[X|Y]$ is orthogonal to $L(Y)$

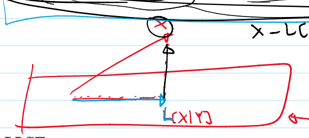
$X - L[X|Y]$ is orthogonal to

$\langle X - L[X|Y], 1\rangle = 0$

$\langle X - L[X|Y], Y\rangle = 0$

$E[X - L[X|Y]] = 0$

$E[(X - L[X|Y])Y] = 0$



$L(Y)$

2. Orthogonal LLSE

(a) Consider zero-mean random variables X, Y, Z such that Y, Z are uncorrelated (i.e they are orthogonal). Show that $L[X \mid Y, Z] = L[X \mid Y] + L[X \mid Z]$.

(b) Explain why for any zero-mean random variables X, Y, Z it holds that:

$$L[X \mid Y, Z] = L[X \mid Y] + L\big[X \mid Z - L[Z \mid Y]\big]$$

$L[X|Y] = E[X] +$

① $L[X|Y] = a + bY$

obeys proj property

a)

X

Δ should be $\perp$ to all elems in $L(Y, Z)$

$X - L[X|Y] - L[X|Z]$

$L[X|Y, Z]$

$L(Y, Z) = $ span$(1, Y, Z)$

WTS
$\langle \Delta, 1\rangle = 0$
$\langle \Delta, Y\rangle = 0$
$\langle \Delta, Z\rangle = 0$

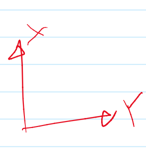
$\Rightarrow L[X|Y] - L[X|Z] = L[X|Y, Z]$

WTS
i) $E[\Delta \cdot 1] = 0$
ii) $E[\Delta \cdot Y] = 0$
iii) $E[\Delta \cdot Z] = 0$

i) $E[X - L[X|Y] - L[X|Z]]$

$= E[X] - E[L(X|Y)] - E[L[X|Z]]$

indep $\Rightarrow$ uncorrelated



i) $E[\Delta]=0$

ii) $E[\Delta Y]=0$

iii) $E[\Delta \cdot Z]=0$

i) $E[X - L[X|Y] - L[X|Z]]$

$= E[X]=0 - E[L[X|Y]] - E[L[X|Z]]$

$= 0 - 0 - 0$ from def of $L[X|Y]$ & b/c $E[\cdot]=0$

indep $\Rightarrow$ uncorrelated $\Downarrow$

uncorr / ortho

ii) $E[(X - L[X|Y] - L[X|Z])Y]$

$= E[((\cancel{X} - L[X|Y]))Y] - E[(L[X|Z])Y]$ iterated exp

$= 0 - 0$

linear fnc of Z

$L[X|Z] = a+bZ$

$\langle Z, Y\rangle = 0$

$E[(a+bZ)Y]$

$= E[aY] + bE(ZY) = 0$

$= 0 +$

& Y & Z are uncorrelated $\Leftrightarrow$

& Y is zero mean

i.e Y & 1 are uncorrelated

$\Rightarrow$ Y is uncorr w/ span$(Z, 1)$

$\Rightarrow$ 0 inner prod

3. Basic Properties of Jointly Gaussian Random Variables

Let $X = (X_1, \ldots, X_n)$ be a collection (or vector) of jointly Gaussian random variables. This means that their joint density is given by

$$f(x) = \frac{1}{\sqrt{(2\pi)^n \det(\Sigma)}} \exp\Big(-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)\Big),$$

for $x \in \mathbb{R}^n$ where $\mu \in \mathbb{R}^n$ is the mean vector and $\Sigma \in \mathbb{R}^{n\times n}$ is the covariance matrix.

(a) Show that $X_1, \ldots, X_n$ are independent if and only if they are pairwise uncorrelated. In your proof, you can assume WLOG that all X_i are zero-mean to make the notation simpler.

(b) Show that any linear combination of $X_1, ..., X_n$ will also be a Gaussian random variable. i.e. show that $Y = a^T X$ is normally distributed for all $a \in \mathbb{R}^n$.
Hint: You can use the fact that the MGF of a jointly Gaussian random vector X is $M_X(t) = \exp(t^T\mu + \frac{1}{2}t^T\Sigma t)$, for $t \in \mathbb{R}^n$.

Linear Estimation:

$$\mathbb{L}[X|Y] := \arg\min_{\text{linear } \hat{X}} \mathbb{E}\left[|X - \hat{X}(Y)|^2\right]$$

after solving for a, b_i $=$ $L[X|Y]$ def I_c

where linear estimators $\hat{X}(Y)$ of form:

$$\hat{X}(Y) = a + \sum_{i=1}^{n} b_i Y_i \quad , \quad a, b_1, \dots b_n \in \mathbb{R}.$$

Q: how to solve for $\mathbb{L}[X|Y]$?

First approach:

$$\underset{a, b_1, \dots b_n}{\text{minimize}} \ \mathbb{E}\left[|X - (a + \sum b_i Y_i)|^2\right]$$

$L[X|Y]$

calculus approach:

$$J(a, b_1, \dots b_n) := \mathbb{E}\left[|X - (a + \sum b_i Y_i)|^2\right]$$

$$= \mathbb{E}|X|^2 - 2a\mathbb{E}X - 2\sum b_i \mathbb{E}XY_i + a^2 + 2a\sum b_i \mathbb{E}Y_i + \sum b_i^2 \mathbb{E}Y_i^2 + \sum b_i b_j \mathbb{E}[Y_i Y_j]$$

$$\frac{\partial J(a, b_1, \dots, b_n)}{\partial a} = 0$$

$$\forall i \quad \frac{\partial J(a, b_1, \dots, b_n)}{\partial b_i} = 0$$

lec 22 / 21